Chapter 3

Directed evolution strategies for enantiocomplementary haloalkane dehalogenases: from chemical waste to enantiopure building blocks

J.G.E. van Leeuwen, H.J. Wijma, R.J. Floor, J.M. van der Laan, D.B. Janssen
Chembiochem 2012 **13**: 137-148.

Abstract

Using directed evolution, we obtained enantiocomplementary haloalkane dehalogenase variants that convert the toxic waste compound 1,2,3-trichloropropane (TCP) to highly enantioenriched (*R*)- or (*S*)-2,3-dichloro-1-propanol, which can be easily converted to optically active epichlorohydrin and therefore are attractive intermediates for the synthesis of enantiopure fine-chemicals. A dehalogenase with improved catalytic activity but very low enantioselectivity was used as the starting point. A strategy that made optimal use of the limited capacity of the screening assay, which was based on chiral gas chromatography, was developed. We used pair-wise site-saturation mutagenesis (SSM) of all sixteen non-catalytic active site residues during the initial two rounds of evolution. The resulting best *R*- and *S*-enantioselective variants were further improved in two rounds of site-restricted mutagenesis (SRM), with incorporation of carefully selected sets of amino acids at a larger number of positions, including sites that are more distant from the active site. Finally, the most promising mutations and positions were promoted to a combinatorial library using a multi-site mutagenesis protocol with restricted codon sets. To guide the design of partly degenerate codon sets for these restricted libraries, we employed structural information, results of multiple sequence alignments, and knowledge from earlier rounds. After five rounds of evolution with screening of only 5,500 clones, we obtained two strongly diverged haloalkane dehalogenase variants that give access to (*R*)-epichlorohydrin with 90% e.e. and (*S*)-epichlorohydrin with 97% e.e., and that contain 13 and 17 mutations around the active site.

Introduction

Directed evolution is a valuable tool for tailoring of naturally occurring enzymes for specific biocatalytic applications. In contrast with rational protein redesign, which is directed towards structure-based construction and testing of small numbers of enzyme variants, directed evolution focuses on the generation of libraries with large genetic diversity and subsequent discovery of improved variants by selection or screening.[1] Iterative cycles of mutagenesis and/or recombination of beneficial mutations, with selection

or screening for improved properties, can produce highly useful biocatalysts.[2-4] Various enzyme properties such as thermostability,[2] resistance to organic solvents,[2] pH optimum,[2] substrate scope,[3] and enantioselectivity[1c,d,4] have been successfully targeted with directed evolution. New tools and strategies directed towards reduction of the experimental work required for a successful directed evolution project are continuously being developed. One important goal is to minimize the screening effort that is required to obtain a desired enzyme by designing libraries with high abundances of improved variants.[1d,5]

Established methods to introduce genetic diversity in directed evolution libraries include various implementations of mutagenic PCR, oligonucleotide-assisted mutagenesis, and in vitro recombination under mutagenic conditions.[6] Sequence saturation mutagenesis (SeSaM)[10] at random positions may also be used. If it is assumed that most mutations are deleterious or neutral, construction of mutant libraries by random methods such as error-prone PCR and random saturation mutagenesis is inefficient when the enzyme properties that it is desired to improve are mainly determined in a small part of the enzyme, such as the active site. To focus saturation mutagenesis at promising regions and to obtain libraries with high abundances of desired variants, methods such as iterative CASTing (combinatorial active site test) and ISM (iterative saturation mutagenesis) were explored by Reetz and co-workers.[7] Structural information is used to select residues that shape the active site and that are expected to influence substrate scope and enantioselectivity. The selected target positions are subjected to mutagenesis in a systematic fashion and this process is repeated until variants with the desired properties are obtained.

In oligonucleotide-assisted mutagenesis, it is also desirable to develop methods that suppress uneven distribution of substitutions resulting from the fact that some amino acids are specified by larger numbers of different triplets than others. This codon bias effect can be reduced by applying partly degenerate NNS or NNK codons instead of the completely random NNN codon in mutagenic primers.[5d]

In order to make a library that favors exploration of functional sequence space, it is possible to use further restricted codons that specify subsets of

the natural amino acids, also referred to as reduced amino acid alphabets.[12] Amino acid subsets can be chosen such that libraries encompass large differences in physical and chemical properties of amino acids without codon bias. Other options are to select sets of amino acids on the basis of phylogenetic analysis, following the assumption that substitutions that do not occur naturally are likely to be poorly tolerated,[13] or to select reduced sets by use of structural information,[15] or to use a database that includes structural, genetic, and literature data.[14] Such smaller sets of amino acids can be useful for controlling library size and might also increase the abundance of desired phenotypes in a library.

In most directed evolution strategies, beneficial mutations obtained from different libraries are combined. This can be done by gene shuffling,[6b] staggered extension with recombination,[6c] and/or by using iterative cycles of mutagenesis.[7,9] Gene shuffling in the presence of synthetic oligonucleotides can direct recombination to predetermined positions along the protein sequence.[10] In these protocols, mutations discovered in multiple variants rather than just the best variant are included in templates or primers for a further round of mutagenesis and screening. Through this process, functional diversity is maintained at key positions and combinations of mutations that act synergistically might be found in later rounds. Statistical methods can be used to select beneficial mutations for inclusion in the next round of evolution.[17]

In the work described here we have further explored a number of the above considerations to develop a strategy for obtaining enantiocomplementary enzymes for conversion of a toxic waste compound into a useful optically active synthetic intermediate. Enantiocomplementary enzymes catalyze the production or conversion of the two enantiomers of a target compound and thereby can give access to both enantiomers in pure or enriched form.[18] In the case of asymmetric transformation of a prochiral substrate, enantiocomplementary enzymes allow production of either product enantiomer with 100% theoretical yield, whereas a kinetic resolution has a maximum yield of only 50%. The development of enantiocomplementary enzymes has been examined for the production of optically active esters with the aid of Baeyer-Villiger mono-oxygenases[19] and of hydroxyketones with the aid of an aldolase,[20] as well as for asymmetric reduction of ketones

with the aid of enoate-reductases[21] and enantioselective epoxidation of terminal alkenes with the aid of cytochrome P450 BM3.[22] Directed evolution of enantiocomplementary enzymes for the conversion of meso-epoxides has been investigated recently.[23]

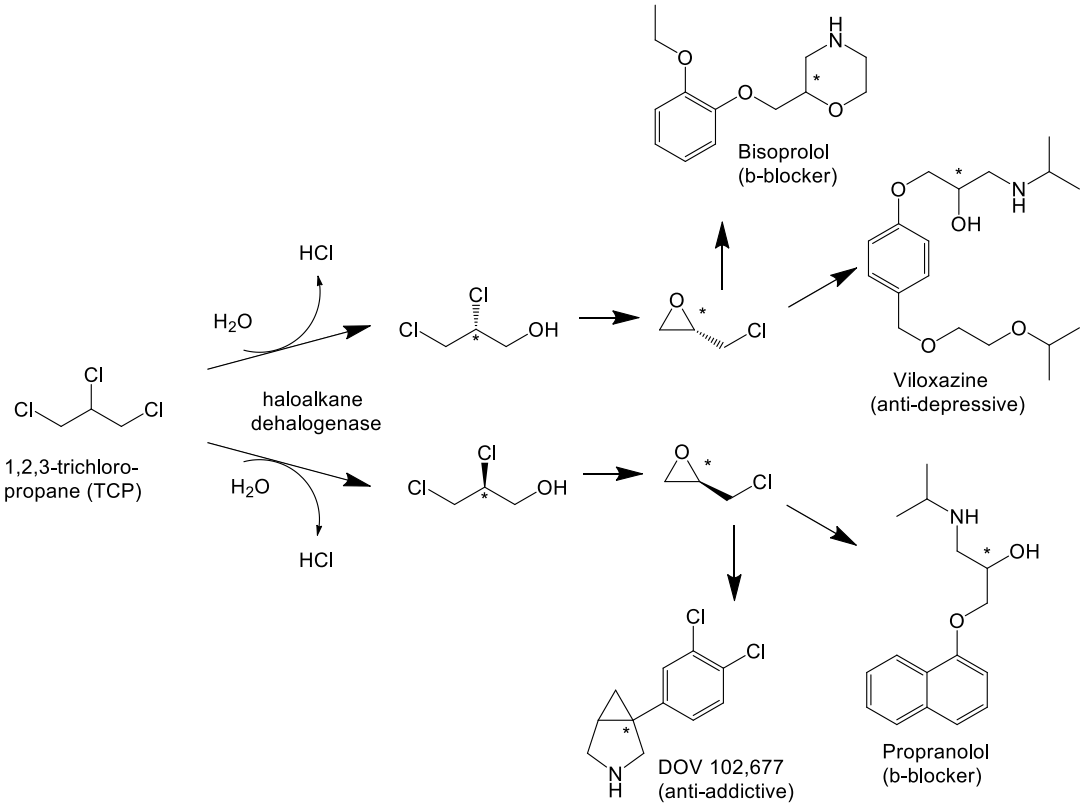

Figure 1. Desired enantioselective conversion of TCP into (*R*)- and (*S*)-2,3-dichloropropan-1-ol by enantiocomplementary haloalkane dehalogenase variants. The products can be converted into epichlorohydrins with inversion of configuration at their chiral centers under basic conditions.[26] Both epichlorohydrin enantiomers can be used as intermediates for the production of chiral bioactive compounds.

The substrate explored here is 1,2,3-trichloropropane (TCP), an industrial waste product that is toxic, extremely recalcitrant to biodegradation, and expensive to dispose of by physical or chemical methods.[24] A significant finding was that this prochiral compound can be slowly converted hydrolytically by haloalkane dehalogenases.[25] The practical feasibility of such an enzymatic transformation would benefit from the product being a useful compound instead of a waste product. We therefore aimed to produce complementary enzymes that would convert TCP into (*R*)- and (*S*)-2,3-dichloropropan-1-ol (DCP), which would in turn give access to the chiral building blocks (*R*)- and (*S*)-epichlorohydrin (Figure 1). Both enantiomers

of epichlorohydrin are valuable fine chemicals that find application in synthetic routes to several pharmaceutical and healthcare products. DOV 102-677, which is under development by Merck for the treatment of alcoholism, for example, is synthesized from (*R*)-epichlorohydrin, whereas (*S*)-epichlorohydrin is a precursor of the nutritional supplement L-carnitine,[26,27] the cholesterolreducing drug atorvastatin,[28] (*R*)-carvedilol, and β-aminoalcohols.[29]

The haloalkane dehalogenases form a subclass of enzymes in the α/β-hydrolase fold superfamily that cleave carbon-halogen bonds.[30] Each is composed of a main domain and a cap domain, with a catalytic pentad located between these domains. The pentad is made up of an Asp-His-Asp catalytic triad and a Trp-Trp or Trp-Asn diad for halide binding. An aspartate residue acts as the nucleophile to displace a halide ion from the substrate. By hydrolysis of the resulting covalent alkylenzyme intermediate, alcohol is released. The only haloalkane dehalogenases with demonstrated activity towards TCP are variants of the enzyme from *Rhodococcus rhodochrous* (DhaA).[25] The wild-type enzyme is widely distributed[31] but has very low activity with TCP.[25a] At least three different directed evolution projects directed towards enhancement of its stability and activity have been carried out.[25a-c] The most active mutant was found by Pavlova *et al.*[25c] and is referred to as DhaA31. It contains five mutations and displays a 27-fold improvement in catalytic efficiency with TCP relative to the wild-type, but with a very low E-value of 1.3 it would allow production of (*R*)-2,3-dichloropropan-1-ol with an e.e. of only 13%. Other haloalkane dehalogenases show large variations in enantioselectivity towards chiral substrates, but the use of prochiral substrates in asymmetric transformations is not well explored.[32]

To obtain enantiocomplementary enzymes for asymmetric conversions of TCP, we set out to accomplish divergent evolution of haloalkane dehalogenase into variants selective for the *R* and the *S* products with use of DhaA31 as the starting point. The modest screening capacity offere-d by chiral gas chromatography prompted us to use strategic considerations carefully in library design throughout the directed evolution process.

Results and Discussion

Design of initial libraries and identification of hot-spots

A number of high-throughput assays for screening of directed evolution libraries for mutants with enhanced enantioselectivity have been developed.[33] However, because only a lowthroughput chiral GC assay was available for the screening of DhaA variants capable of producing enantioenriched (*R*)- or (*S*)-2,3-dichloropropan-1-ol from TCP, it was considered essential to construct libraries containing high frequencies of mutations contributing to the desired phenotypes.

One way to increase the abundance of improved mutants in a library is to focus mutagenesis on positions in the sequence where, on the basis of the 3D structure, mutations would be expected to influence the phenotype that should be modified.[7] Obviously, in cases in which the goal is a change in substrate or product specificity, this is around the active site. A molecular model of DhaA31 containing five mutations (I135F, C176Y, V245F, L246I, Y273F) relative to the wild-type enzyme[25c] was constructed with use of the published X-ray structure for the wild-type enzyme as a template.[34] Inspection of the model with TCP docked in the active site revealed that 16 amino acids not belonging to the catalytic pentad surround the bound TCP (Figure 2). These first shell non-catalytic positions were selected as targets for the first round of mutagenesis directed towards the production of variants for the formation of (*R*)-DCP and of (*S*)-DCP.

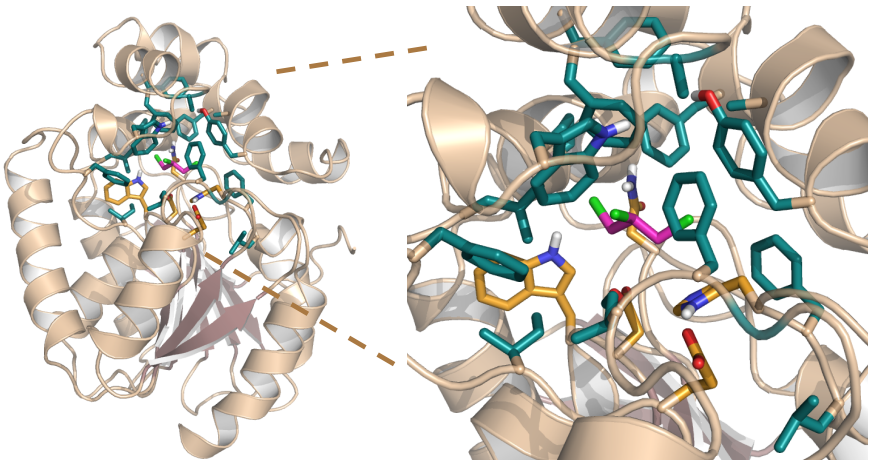

Figure 2. Structural model of DhaA31 with TCP (magenta) docked in the active site. Catalytic residues are displayed in orange and the targeted first shell residues are displayed in turquoise.

Another consideration in library design was the fact that much larger numbers of different enzyme variants can be tested with the same screening effort if repeated sampling of identical clones is avoided. Comprehensive or deep screening of libraries is always accompanied by extensive oversampling (e.g., five-fold to obtain 85% coverage of a two-site saturation library).[35] This oversampling can be avoided through partial screening of larger libraries or by screening multiple libraries in parallel. We thus consider it inefficient to construct smaller libraries just because it is easier to screen them completely. In other words, as long as the abundances of desired mutants are similar, it is preferable to omit genetic diversity at the screening step instead of at the library construction step, because without oversampling a larger amount of sequence space can be explored. The choice for partial screening of multiple pair-wise libraries also has the advantage that possible synergistic effects, which are known to be important in directed evolution, are more intensively explored.[36]

For the first round of directed evolution we decided, on the basis of the structural model with TCP docked in the active site, to use oligonucleotide-assisted saturation mutagenesis for the selected 16 active site positions, combined in 14 partly overlapping pair-wise libraries, which could be created with standard QuikChange mutagenic PCR with use of NNS codon degeneracy. Of the 16 positions, ten are represented in two libraries, allowing thousands of different combinations that could potentially have synergistic effects (Figure 3). The 16 selected positions included sites discovered to be important for TCP dehalogenase activity in earlier experiments.[25a,c] The total number of possible variants in the combined libraries was 6,400. From each library, 96 randomly picked variants were tested for enantioselectivity and catalytic efficiency by chiral GC, and many active clones with increased enantioselectivity were found in different libraries (Figure 4). About 21% of the sequence diversity represented in the pair-wise libraries was tested and about 89% of the tested clones used for chiral GC assays should be unique.

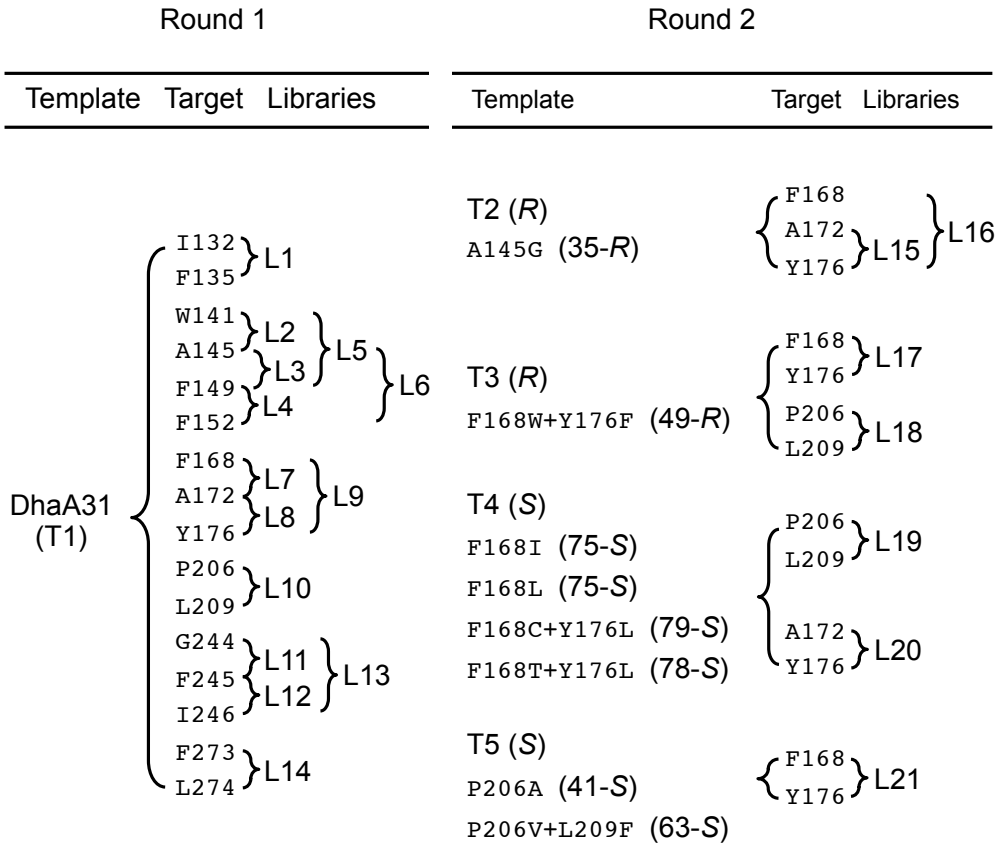

Figure 3. Pair-wise mutant library design strategy. Round 1 used DhaA31 as template (T1); both the 16 target positions selected on basis of the structure and the pair-wise combinations are shown. Round 2 libraries are based on templates (T1, T3, for (*R*)-DCP selectivity), or equimolar mixtures of templates (T4, T5, *S*-selectivity) and promising target positions found in round 1 in new pair-wise combinations. The e.e. values of the templates used for the second generation are indicated in brackets.

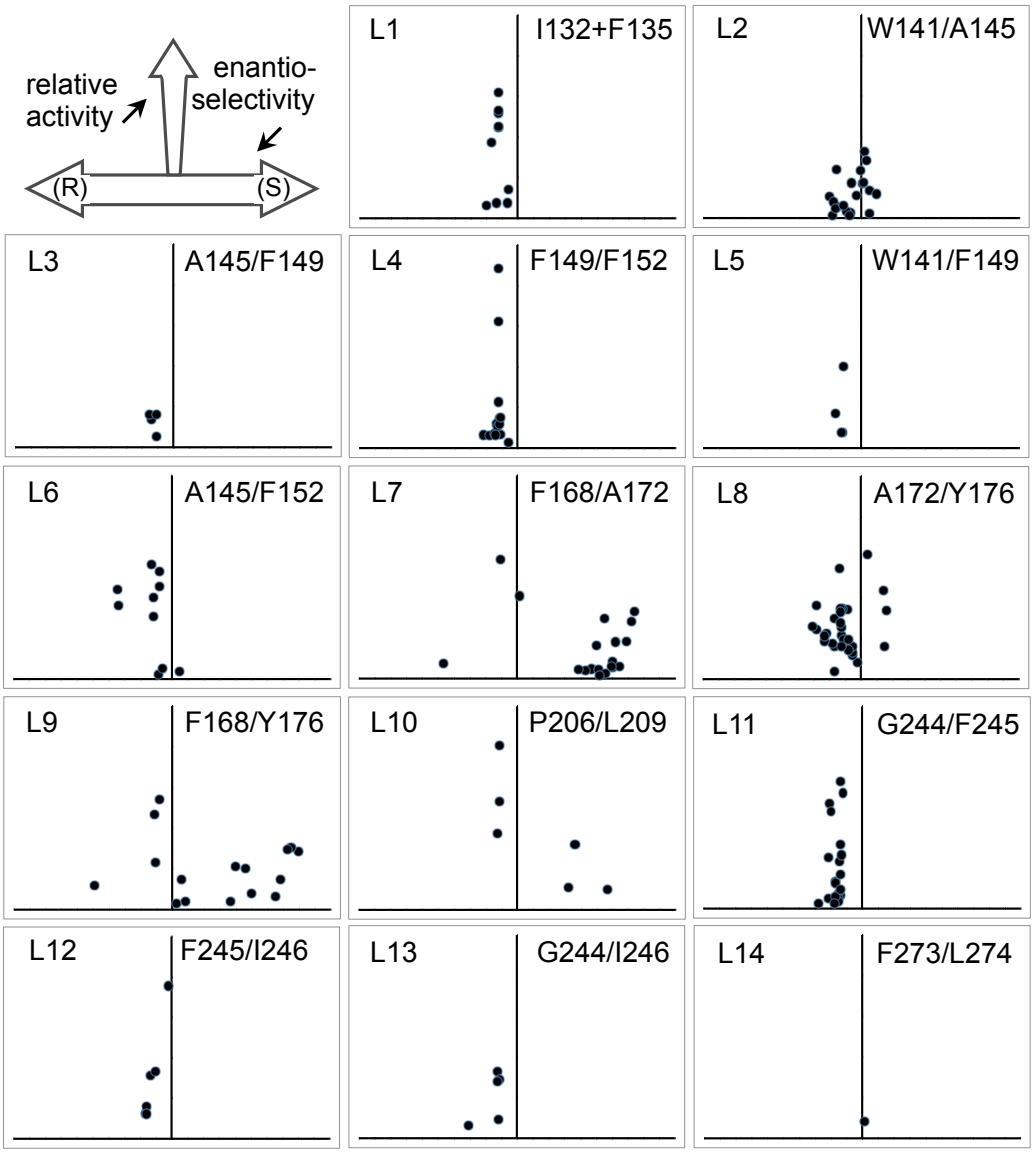

Figure 4. Screening results for 14 first-round mutant libraries. The product e.e. values (from 100% *R* to 100% *S*) are displayed on the horizontal axes and the relative activities (conversion compared to wild-type) are displayed on the vertical axes. Dots indicate individual clones with measurable activity and enantioselectivity.

The results of the first round allowed the identification of positions in the sequence at which mutations, either separately or in combination, introduce significant enantioselectivity in DhaA variants in the conversion of TCP. The results also revealed positions at which mutations are detrimental for activity. In library 14, for example, only 2% of the clones were active, suggesting that most mutations at either position in this library have negative effects on activity (Figure 4). Positions 273 and 274, which were those mutated in library 14, were thus not included in the next round of

mutagenesis. The same holds for positions 132, 135, 149, 245, and 246, which did not yield variants with high enantioselectivity on mutation, even when tested in different combinations. In contrast, mutations at positions 168, 172, and 176 occurred in different libraries in which more than 40% of the mutants were active, and many variants with significant *R* or *S* enantioselectivity were detected in these libraries (L7, L9; Figure 4). These apparent hot-spot positions were promoted to the next round of mutagenesis and screening. Positions 145, 206, and 209, of which the last two had been tested only in a single pair-wise library in round 1 and had also given positive hits, were also promoted to round 2 and were subsequently tested in combinations with other positions. The new combinations were made both by use of templates with substitutions at positions 145, 206, and 209 and primers with mutations at other positions, and by targeting of these positions with new mutagenic primers in combinations with templates substituted at other sites (Table 4). An example of an interesting hot-spot is position F168, located in an α-helix close to the TCP binding site. In the best *R*-selective variant F168 was mutated to a tryptophan (giving (*R*)-DCP with an e.e. of 47%) whereas in the best *S*-selective mutant (78% e.e.) F168 was mutated to a cysteine, indicating that different mutations at the same position can give opposite DCP enantiomers as products.

Intensive exploration of mutations at hot-spots in new combinations

After enantioselectivity had been observed in different DhaA variants and hot-spots had been identified in the first round of evolution (Figure 4), further improvement was sought through more intensive exploration of sequence diversity at these hot-spots. To achieve further divergence of *R* and *S* product enantioselectivity, beneficial mutations and influential sites were retested in new combinations through the construction of a second generation of pair-wise site-saturation libraries. As templates we used not only individual plasmids with mutations but also mixtures of plasmids that carried promising mutations as discovered in round 1 (e.g., a mixture of plasmids with P206A and P206V+L209F was used in the template T5). New combinations of mutations that had exhibited clear effects on enantioselectivity in round 1 (e.g., F168+Y176 in library L21, Figure 3) were tested. In this way, target sites were revisited in round 2 in even

larger numbers of combinations. High diversity was achieved in the libraries, whereas the frequencies of inactive variants were kept low. A total of seven libraries were constructed with 14 new combinations of templates and target positions, giving a total library size of 5,600 possible variants (Figure 3).

To make optimal use of the limited screening capacity of the chiral GC-based method, a prescreening assay for dehalogenase activity was used, allowing removal of inactive variants that might appear due to disruption of the integrity of the catalytic site. In this pH indicator plate assay we used bromoethane, which was selected because it is nonchiral and easily converted by active haloalkane dehalogenases (because the carbon-bromine bond in bromoethane is much easier to break than the carbon-chlorine bond in TCP). Furthermore, bromoethane has a low toxicity and viable cells can readily be recovered from active colonies after exposure. Prescreening on pH indicator plates increased the proportions of TCP-active clones detected in subsequent GC assays to be 60-100%, depending on the templates and primers that were used, whereas without prescreening the proportions of TCP active clones were only 10-50%.

After prescreening and testing of product e.e. values by chiral GC, several mutants with further improved enantioselectivity were indeed found in this second round of directed evolution. The best variants gave (*R*)-2,3-dichloropropan-1-ol with 60% e.e. and *S* product with an e.e. greater than 80%. The sequences of the eight variants that produced (*S*)-2,3-dichloropropan-1-ol with e.e. >80% revealed that these were based on a template with position F168 changed either to leucine (six out of eight) or to isoleucine (two out of eight), whereas the best first-generation clones had contained the mutations F168C or F168T. This shows that nonoptimal mutations can reappear in different contexts in later rounds of evolution, which is in agreement with the occurrence of synergistic effects observed in other enzymes.[36] On the other hand, even though further improved *S*-enantioselective variants (up to 80% product e.e.) were discovered in libraries that included the sites P206 and L209, these positions were discontinued because of the very low catalytic activities in all libraries in which these positions were mutated, even though they were tested in different combinations with other positions. The observations show that

promotion of multiple beneficial mutations to a subsequent round of mutagenesis by use of template mixtures is an efficient strategy for creating libraries with large numbers of different combinations of mutations, which allows intensive exploration of parts of sequence space in which synergetic effects can become manifest.

Revisiting first shell residues in evolved contexts by restricted mutagenesis

The best *R*- and *S*-enantioselective variants that were obtained after two rounds of evolution are all modified in an α-helix that lines the active site (Figure 5). These were used as new templates for further divergence. Improvement was sought by mutagenesis of first shell positions that had displayed modest influences on enantioselectivity in the first round.

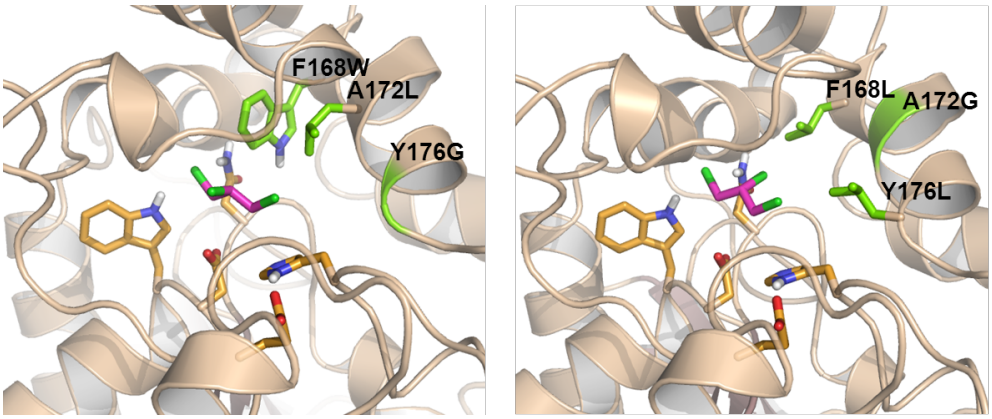

Figure 5. Close-ups of the active sites of the best clones r2-60R (left) and r2-85S (right), after two rounds of evolution. Catalytic residues are in orange, mutated residues in green. Left: Best *R*-enantioselective variant with three mutations (F168W+A172L+ Y176G). TCP is docked in the active site in a configuration that leads to l-2,3-dichloropropan-1-ol formation. Right: Best *S*-enantioselective variant with different mutations at the same positions (F168L+A172G+Y176L) and with TCP bound in a configuration that leads to *S* product formation. The positions 168, 172, and 176 have a strong steric effect on the bound substrate in the active site. For modeling see the Experimental Section.

Because the earlier rounds had identified substitutions that were not accepted in active variants, or only with low frequencies, it was decided not to apply saturation mutagenesis, but to introduce a specific restricted set of amino acids at each target position (site-restricted mutagenesis, SRM). Furthermore, rare codons, stop codons, and amino acid substitutions that were expected on the basis of their positions in the structure to influence enzyme production or catalytic performance negatively (or to abolish them entirely) were omitted when possible.

The restricted sets of amino acids were introduced with the QuikChange oligonucleotide-assisted mutagenesis approach with use of PCR primers that included partly undefined codons at target sites (Table 1). At position 149, where only wild-type phenylalanines appeared in first- and second-round hits, for example, substitution was restricted to Phe and Tyr, which was achieved with a partly undefined TWC codon (W stands for A+T). Similarly, because multiple sequence alignments of haloalkane dehalogenases indicated restricted variation at positions 157 and 242, these were also tested with partly undefined codons, with introduction of I157(I,L,F,M) and of T242(T,S,N), respectively. Furthermore, some mutations that had been encountered frequently – F246C and I246L, for example - indicated that hydrophobic amino acids were preferred at certain positions in improved variants, so these sites were targeted with oligonucleotides encompassing partly undefined codons that specifically cover hydrophobic residues. In this way, five SRM libraries were constructed, each targeting four sites and each contains approximately 280 possible variants.

Table 1. Example of a site-restricted mutagenesis (SRM) library based on the use of partly undefined (ambiguous) codons.

target position	W141 [a]	A145	F149	F152
desired set of amino acids	A,C,F,G,L,V,W,Y	A,G,V	F,Y	C,F,L,W,Y
restricted codon primer 1	TDS [c]	GBC [b]	TWC [b]	TDS [c]
encoded amino acids	C,F,L,W,Y,*	A,G,V	F,Y	C,F,L,W,Y,*
restricted codon primer 2	GBC [b]	-	-	-
encoded amino acids	A,G,V	-	-	-

[a] Two mutagenic primer pairs were used in a 2:1 ratio to achieve the desired set of substitutions at position 141.

[b] Code for partly degenerate codons: R = A+G; Y = C+T; M = A+C; K = G+T; S = C+G; W = A+T; B = C+G+T; D = A+G+T; H = A+C+T; V = A+C+G; N = A+C+G+T.

[c] With the TDS codon degeneracy a stop codon (*) is also included, but pre-screening will discard clones with this mutation.

Obviously, most defined subsets of amino acids (there are 1,048,555 possible combinations of two to 20 natural amino acids at a single site in a mutant library) cannot be covered with a single partly undefined codon. Therefore, in several cases mixtures of primers with partly undefined codons were used during oligonucleotide-assisted mutagenesis, an example of which is given in Table 1 for the W141(A,C,F,G,L,V,W,Y) restricted mutagenesis. To facilitate the search for the best defined (mixtures of) partly undefined codons that can be used for efficient construction of these SRM libraries, search methods have been developed.[7c,16] We developed a codon search algorithm (CoFinder) capable of finding both a suitable partly undefined codon as well as mixtures of partly undefined codons, with minimal codon bias and optimal avoidance of undesired mutations in the library (Chapter 4). The use of primer mixtures with a different partly undefined codon at the same position allows more precise coverage of a desired set of amino acids than a single primer.

Screening of the SRM libraries designed to increase *R*-enantioselectivity resulted in several hits, of which the best variant produced (*R*)-2,3-dichloropropan-1-ol with an e.e. that had been improved from 60 to 75% (Table 2). The catalytic activity of this third-generation variant was significantly enhanced relative to its parent clone. The substitutions responsible for this improvement are F245Y and I246L. These are at positions that were also targeted in library 12 during the first round of evolution, even though hardly any variants displaying altered enantioselectivity had been observed in that library (Figure 4). This again suggests that the effects of mutations are strongly context-dependent and/or that functional sequence space was more efficiently probed by SRM than by site-saturation mutagenesis, because the frequency of beneficial variants is elevated.

In a library directed towards further enhancement of *S*-enantioselectivity, two variants that produced (*S*)-2,3-dichloropropan-1-ol with e.e.s of 90% were discovered, in relation to 80% for the best round 2 variants. Also in these variants, I246 was substituted with a leucine, hence representing reversion of a DhaA31 mutation to the amino acid carried by the wild-type DhaA at this site.

Table 2. Mutations in selected *R*- and *S*-enantioselective DhaA variants obtained in subsequent rounds of enantiodivergent directed evolution.

Structure [a]		S	L	L	L	L	L	L	H	H	H	H	H	H	H	H	L	H	H	H	H	L	L	L	L	L	L	H
		C	F	I	R	F	P	E	W	Q	A	F	A	L	K	Y	V	L	W	F	V	G	T	F	I	I	P	E
Round	e.e.	128	131	132	133	135	136	140	141	165	167	168	172	173	175	176	178	202	203	205	219	241	242	245	246	247	248	251
(*R*)-selectivity [b]																												
r1	47 (*R*)											W				F												
r2	60 (*R*)											W	L			G												
r3	75 (*R*)											W	L			G								Y	L			
r4	84 (*R*)					L	T					W	G			L								Y	L			
r5	90 (*R*)					L	T				M	W	L		A	G							N		L		T	
(*S*)-selectivity [b]																												
r1	78 (*S*)											C				L												
r2	85 (*S*)											L	G			L												
r3	90 (*S*)											L	G			L							S		L			
r4	93 (*S*)	F	I	V	A							L	G			L							S					
r5	97 (*S*)	F	M	V				D	F	R		L	G	L				I	L	W	L	A		F	L	V		Y

[a] S =sheet, L = loop, H = helix. All mutations are in a DhaA31 background.
[b] In the first column r# indicates the round number from which the mutant is obtained.

Although this revisiting of the first shell sites in an improved context by SRM resulted in the discovery of a few further improved *R*- and *S*-enantioselective variants, screening of most third-round libraries did not deliver mutants with enhanced enantioselectivities, which suggests that the remaining first shell positions that were revisited in round 3 SRM libraries either do not contribute to enhanced enantioselectivity or do so only in rare combinations with other amino acid substitutions. Further divergence of the enzymes was therefore sought at more distant sites.

Exploring distant residues with site-restricted mutagenesis (SRM)

Engineering of enzyme enantioselectivity is often focused on first shell residues around the substrate binding pocket, but mutations at more distant positions can also have effects, either through indirect interactions or by influencing protein dynamics.[37] Further mutations that might enhance enantioselectivity of DhaA variants were thus examined at a large number of distant sites. A total of 52 positions located close to the substrate binding pocket, either in the cap domain or in the top part of the main domain, were selected on the basis of structural inspection and ph ylogenetic analysis. Keeping in mind that the modest screening capacity required the construction of libraries with large representation of beneficial variants, we used another round of site-restricted mutagenesis (ISRM, iterative site-restricted mutagenesis). The selection of reduced sets of amino acids for all target sites was inspired by the natural diversity that was analyzed with the web tool Hot-Spot Wizzard.[8,13] C128, which is located at the top part of the main domain, for example, was mutated only to Cys or Phe, whereas F131, positioned in a loop region connecting the main domain with the cap domain, was mutated only to Phe, Leu, Ile, and Met, because phylogenetic data suggested a strong preference for hydrophobic residues at these sites.

With the help of CoFinder, primers for nine site-restricted libraries were designed for further enhancement of *R*-selectivity in round 4, with use of isolated plasmid DNAs of the best *R*-selective hits of round 3 as templates. For further divergence in the direction of *S*-enantioselectivity, thirteen libraries were constructed with use of an equimolar mixture of plasmid DNAs from the best *S*-selective variants available as templates. Four to five sites were addressed in each library, and the libraries were again screened in parallel.

Prescreening showed that the frequencies of active clones in these round 4 libraries were fairly high (up to about 70%), indicating either that the phylogeny- and structure-inspired site-restricted mutagenesis indeed reduced the frequency of dysfunctional mutants or that these distant positions were less critical to enzyme activity. Several hits that displayed further enhanced *R*- or *S*-enantioselectivity were discovered. In a library targeting *R*-enantioselectivity, a clone that produced 2,3-dichloropropan-1-ol with an e.e. of 84% was found, in relation to 75% e.e. for the best variant in the previous round. In another library, targeting production of (*S*)-2,3-dichloropropan-1-ol, *S*-enantioselectivity was improved from 90 to 93% product e.e.. A shift in the focus of mutagenesis from the first shell to more distant locations thus indeed identified new positions that influenced enantioselectivity and uncovered substitutions that gave further improvement.

Final recombination of sites in consensus-based SRM

To improve the *R*- and *S*-enantioselectivities in the diverging DhaA variants further, a fifth round of directed evolution was conducted. All important potentially contributing substitutions were now combined in a single library with simultaneous exploration of new genetic diversity through their occurrence in different combinations. To reduce randomness during combination, we used a site-restricted mutagenesis protocol that allows combinatorial incorporation of defined sets of substitutions at multiple positions. The positions were ordered in short regions that could each be covered either with a single mutagenic oligonucleotide or with mixture of such mutagenic oligonucleotides carrying several partly undefined codons (Figure 6).

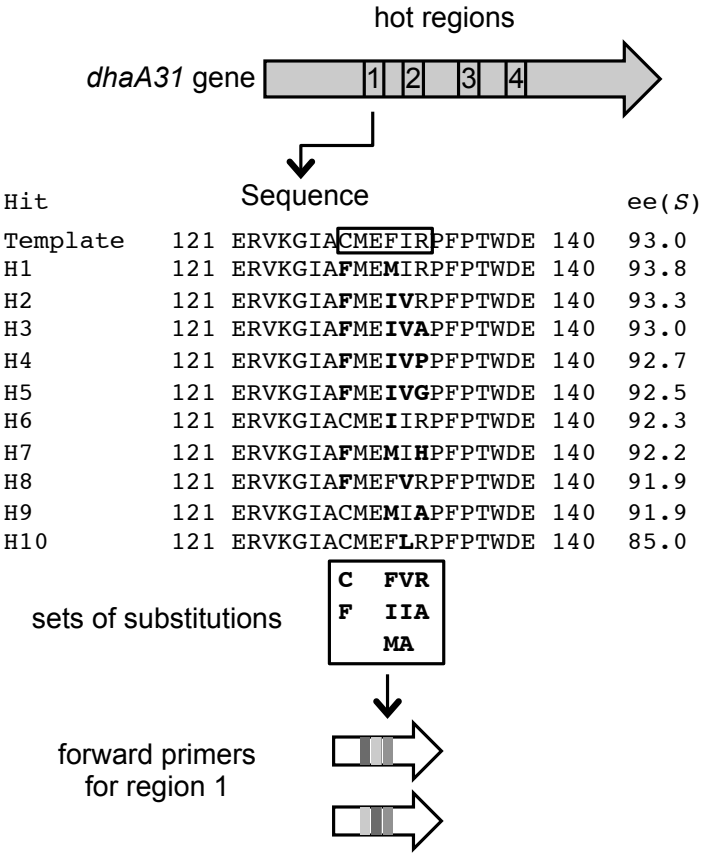

Figure 6. Combinatorial library design for the final round of directed evolution for (*S*)-product-selective haloalkane dehalogenase. Four hot regions, each encompassing four to six influential positions, were subjected to site-restricted mutagenesis (SRM). For each position, a restricted set of amino acids was selected on the basis of mutants from previous rounds. Ten such earlier mutants (H1-H10) are given here, together with their e.e. values. The restricted sets of amino acids were introduced by QuikChange PCR with use of oligonucleotides or oligonucleotide mixtures containing partly undefined codons at the target positions. The first hot region, with segment C128-R133 (boxed) for example, is modified to (C/F)ME(F/I/M)(V/I/A)(R/A). The template sequence is the best hit of round 4, (r4-93S, Table 2) and is identical to DhaA31 in the region that is shown.

To select the influential positions and the preferred potentially beneficial substitutions, 70 hits with improved enantioselectivity, plus some variants with lower enantioselectivity from the previous round, were sequenced. The identities of beneficial mutations and their combined occurrences were determined by multiple sequence alignments and used in the design of the round 5 libraries. These were constructed by consecutive rounds of QuikChange PCR reactions, which started with separate template mixtures prepared from earlier hits that were *S*- or *R*-enantioselective. For example, the template used in round 5 for further enhancing *R*-enantioselectivity

consisted of a mixture of four plasmids from round 4, each carrying seven or eight mutations. Restricted mutagenic oligonucleotides were used in a combinatorial library design such that they covered different regions along the protein sequence. To avoid accumulation of deleterious combinations of mutations, in each QuikChange round one fifth of the template DNA was omitted from the PCR reaction and added to the obtained PCR product before the subsequent QuikChange round that targeted the next hot region was performed. By this approach, two different combinatorial libraries for enhanced *R*- and *S*-enantioselectivity were constructed.

The combinatorial libraries used for *S*-selectivity in round 5 again contained very high frequencies of active clones (about 80%), and this time almost 90% displayed better *S*-enantioselectivity than the best earlier variant. This clearly shows that the previous rounds of evolution and structural inspection had indeed uncovered clusters in sequence space from which further mutants could successfully be selected by testing new combinations of mutations and using low-throughput chiral GC screening. The best *S*-enantioselective variant produced (*S*)-DCP with 97% e.e. (Figure 7, Table 2). A similar improvement was observed with the constructed *R*-selective library, which produced a clone that yielded (*R*)-DCP with 90% e.e., significantly better than the 84% e.e. of the best round 4 variant that was found.

The best *R*- and *S*-selective hits that were found in the fifth and final round of directed evolution carried 13 and 17 amino acid substitutions, respectively. Most of these are located around the substrate binding pocket (Table 2, Figure 7). Some mutations were found further away in the cap domain and in the region where the main domain and cap domain make contact. Interestingly, several highly contributing substitutions appearing in this round were present in loop regions in the top part of the main domain and in a loop that connects the main domain with the cap domain. The substitutions F135L+P136T in the best *R*-selective variant and the substitutions E140D+ W141F in the best *S*-selective variant are located in this loop, for example. Mutations in the connecting loop were also observed in a laboratory evolution experiment with the phylogenetically and structurally related *Xanthobacter autotrophicus* haloalkane dehalogenase.[39] Most substitutions in the active sites of the best *R*- and *S*-enantioselective variants were located at identical positions. However, later substitutions appearing at

regions more distant from the active site were found at different positions in the *R*- and *S*-selective mutants (Table 2).

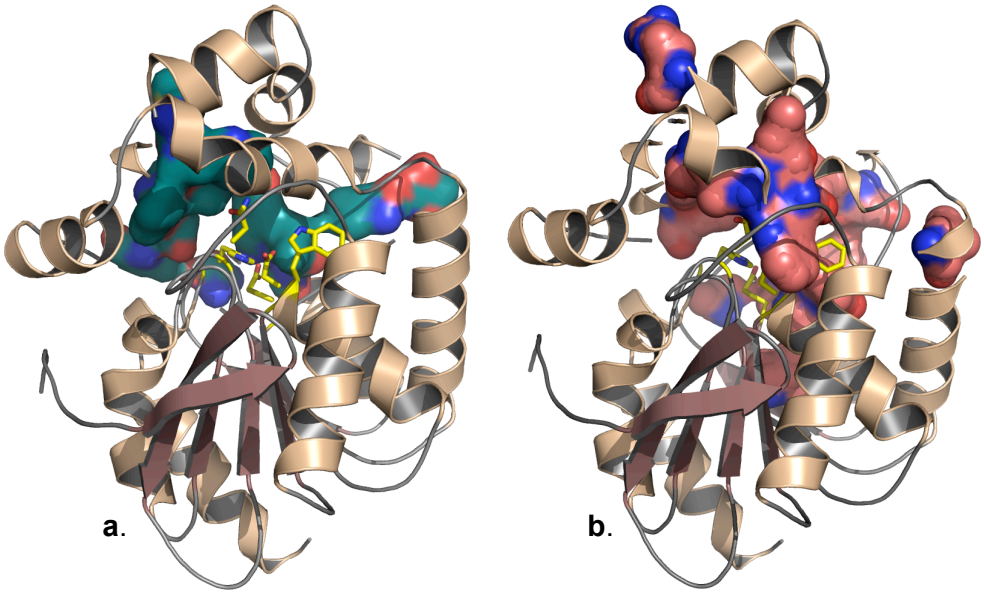

Figure 7. Structural models of the final best enantiocomplementary DhaA variants r5-90R and r5-97S. **a**. Variant with 13 mutations producing (*R*)-DCP. **b**. Variant with 17 mutations producing (*S*)-DCP. The models show the mutated amino acids in surface display revealing the wide distribution. Catalytic residues are shown in yellow.

Trajectory of divergent evolution

In five rounds of directed evolution, the enantioselectivity of the DhaA variants diverged from 13% *R*-product enantioselectivity in the initial DhaA31 variant to 90% e.e. *R*-product selectivity for the DhaA variant r5-90R and 97% *S*-enantioselectivity for DhaA-r5-97S (Figure 8, Table 3).

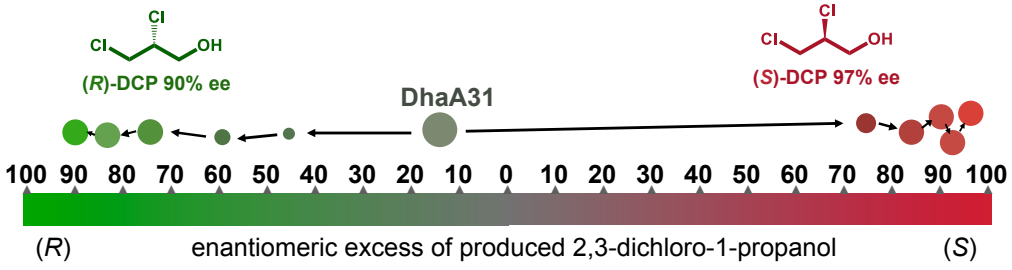

Figure 8. Schematic overview of the obtained evolutionary pathway.

Table 3. Contribution of the evolutionary rounds to the differential transition state energies of the dehalogenase reactions with the two pro-enantiomers.

Round and mutant	(*R*) selec-tivity	$\Delta_{R\text{-}S}\Delta G‡$ (kJ.mol^{-1})	contri-buion (%)	round and mutant	(*S*) selec-t-ivity	$\Delta_{R\text{-}S}\Delta G‡$ (kJ.mol^{-1})	Contri-bution (%)
DhaA31	13 (*R*)	0.62	-	DhaA31	-13 (*S*)	-0.62	-
r1-47R	47 (*R*)	2.63	28.8	r1-75S	75 (*S*)	5.02	49.4
r2-60R	60 (*R*)	3.57	13.5	r2-85S	85 (*S*)	6.48	12.8
r3-75R	75 (*R*)	5.02	20.7	r3-90S	90 (*S*)	7.59	9.8
r4-84R	84 (*R*)	6.30	18.4	r4-93S	93 (*S*)	8.55	8.4
r5-90R	90 (*R*)	7.59	18.6	r5-97S	97 (*S*)	10.8	19.6

The differences in transition state energies were calculated with Equation 1. The reported percentages are the additional contributions that the new mutations in every round make to the transition state energy difference in the final variants. Mutations in the best variants are given in Table 2.

To determine the effects of individual evolutionary rounds on the enantioselectivity we calculated the contributions of the individual steps to the differences in transition state energy between the two reactions by use of Equation 1. In this equation, $\Delta_{R\text{-}S}\,\Delta G‡$ is the difference in transition state energy in comparison of conversion into (*R*)-DCP and into (*S*)-DCP, e.e. has its usual meaning (0-1 scale), R is the gas constant, and T is the absolute temperature.

$$\Delta_{R-S}\,\Delta G^{‡} = ln\frac{1+e.e.}{1-e.e.} \times RT \qquad \text{Equation 1.}$$

The enantiodiscrimination must be based on differences in binding energy and/or reactivity between the pro-*S* and pro-*R* binding modes of TCP in the active site. This concept makes the high enantioselectivity obtained in the mutants really impressive, because the reaction does not take place at the prochiral carbon atom and the enantiodivergent enzyme variants have to discriminate between a hydrogen and a chlorine substituent to bind the substrate in a mode that fits the appropriate product enantiopreference.

Equation 1, can be used to analyze evolutionary landscapes.[38] Both for *R*- and for *S*-enantioselectivity, the initial round contributed most to the evolution of the transition state energy differences that cause enantioselectivity (29 and 49%, respectively). The two subsequent rounds, which targeted first shell residues in an improved context, contributed 34 and 23% to the final *R*- and *S*-enantioselectivities. The last two rounds, in which

distant positions were addressed, contributed 37 and 28% to the free energy differences for enhancing *R*- and *S*-selectivities. These results indicate that it is not only positions close to the active site that are attractive targets for mutagenesis, but that positions more distant from the active site can also contribute considerably to directed evolution of enantiocomplementary enzymes.

Catalytic properties of the best R- and S-enantioselective variants

The best *R*- and *S*-enantioselective mutants that we found (Table 3) were purified by His-tag Ni-agarose chromatography. Both variants were >95% pure by SDS-PAGE and the kinetic properties for TCP conversion were determined (Table 4).

Table 4. Kinetic properties of divergently evolved DhaA variants.

Variant	e.e.	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
DhaA31	13 (*R*)	1.26	1.2	1050
DhaA wild-type	13 (*R*)	0.08	2.2	36
r5-90R	90 (*R*)	0.16	6.5	25
r5-97S	97 (*S*)	0.18	0.65	277

The k_{cat} values of both variants were lower than the k_{cat} of DhaA31, but approximately twice as high as the k_{cat} of the wild-type DhaA from which DhaA31 is evolved.[25d] The K_M value of the best *R*-selective clone was three times higher than that of the wild-type DhaA, whereas the K_M value of the best *S*-selective variant was three times lower. The catalytic efficiencies of both enantioselective variants were lower than the k_{cat}/K_M value of the starting point DhaA31 (Table 4). Of the five amino acid substitutions that distinguish DhaA31 from wild-type DhaA, only one and two substitutions remained unaltered in the two best *R*- and *S*-enantioselective variants, respectively.

The stabilities of the best divergently evolved variants were also tested. The wild-type DhaA displayed an apparent T_m of 49°C as determined by thermofluor measurements and circular dichroism. The corresponding values for the mutants were determined by the thermofluor method and were 55, 51, and 57°C for DhaA31, DhaA31-90R and DhaA31-97S, respectively, indicating that stability was hardly affected in either of the highly mutated enzymes.

Conclusions

Directed evolution for obtaining improved biocatalysts often requires time-consuming and expensive experimentation involving multiple rounds of mutagenesis and screening. Although various powerful approaches have been developed, they might not be applicable in many cases. Especially if protein production procedures take very long, or if the target property cannot be screened by high-throughput procedures, methods that depend on testing large numbers of mutants or many rounds of mutagenesis and screening will not be successful. In this investigation we have developed strategies to design efficient libraries that reduce screening efforts. A number of strategic considerations was incorporated in the approach that we followed for directed evolution of enantiodivergent TCP-converting haloalkane dehalogenases.

Firstly, unlike in most common protocols, we constructed libraries that encompassed multiple combinations of pair-wise mutations around the active site. Obviously, this requires the availability of structural information, for which we used a homology model built with a sequence containing five substitutions relative to the wild-type template. The resulting libraries were screened only partially, and nowhere was any attempt made to reduce library size, which would be required for complete or deep screening of the libraries, because the same screening effort examines a much larger fraction of sequence space when a large library (in our case encompassing multiple combinations of mutations) is screened partially, in comparison with the more or less comprehensive screening of a small library (which typically requires four- to six-fold oversampling for a two-site saturation library for 90-95% coverage).

Secondly, to maintain functional genetic diversity during evolution, we proceeded to next-generation libraries with multiple mutations and target sites from earlier hits by using both template mixtures and primer mixtures in mutagenic PCR reactions for library preparation. In this way, new mutations were introduced in various combinations at two to five positions at the same time, allowing diverse synergetic effects in the resulting libraries, and increasing the chances of discovery of beneficial variants that are dependent on such synergetic effects. In other words, individual rounds

were more focused on identifying influential positions than on the best mutations, because these are very context-dependent anyway.

Thirdly, to increase the frequencies of positive variants in mutant libraries, we restricted substitutions at target positions to selected functional subsets of amino acids. These subsets were identified by sequence data from earlier rounds, inspection of the three-dimensional structure of the enzyme, and phylogenetic analysis of homologous haloalkane dehalogenases. The resulting site-restricted mutagenesis (SRM) libraries were constructed with restricted ambiguous codon sets, which were conveniently selected with a codon search algorithm (Chapter 4). To optimize coverage of any desired set of amino acids at a specific position in a library, CoFinder will suggest mixtures of primers, each with a specific partly undefined (ambiguous codon), that are used in the same mutagenic PCR reactions.

With these strategies we obtained enantiocomplementary haloalkane dehalogenase variants that catalyze the asymmetric conversion of TCP, a highly toxic industrial waste compound. The structure of TCP means that it is a very challenging target, because any enantioselectivity is caused by different binding modes of this prochiral substrate in the transition state even though TCP offers few interactions that restrict its active site dynamics. The highly diverged variants that were obtained after five rounds of evolution differ at 25 positions, yet were discovered by chiral GC screening of just 5,500 clones. The best *R*- and *S*-enantioselective mutants carry 13 and 17 mutations, respectively, relative to DhaA31, and produce the chiral building blocks (*R*)-2,3-dichloropropan-1-ol with 90% e.e. and (*S*)-2,3-dichloropropan-1-ol with 97% e.e.. The approach followed here thus appears to be an attractive strategy for directed evolution of enantiodivergent enzymes for the conversion of prochiral TCP. The approach might be also be useful for other enzymes in which catalytic properties are governed by subtle differences in substrate positioning, especially if high-throughput enzyme production or screening assays are unavailable.

Materials and Methods

Molecular Modeling

To predict the structure of mutated proteins, a computational protocol was developed under YASARA (www.yasara.org).[41] The protocol consists of six rounds of optimization. In each round, the mutated side-chains are energy optimized in implicit water by sampling discrete side-chain conformations (rotamers). This is followed by a gradient energy minimization of the mutated side-chains and the nearby protein in explicit water. Prior to predicting the mutant DhaA structures, TCP was docked in pro-*R* and pro-*S* orientations suitable for catalysis into the template PDB file 1BN6[33,42] using Autodock 4.0 as described earlier.[43,32b] Figures were prepared by the use of Pymol software (www.pymol.org).

Cultivation and expression

The dhaA31 gene[27c] was kindly provided by Prof. Jiri Damborsky (Masaryk University, Brno, Czech Republic). The gene was cloned with an N-terminal His_6 tag in the pBAD/Myc-His expression vector with use of NdeI and XhoI restriction sites.[31b] The construct was transformed to and produced in *E. coli* XL10-Gold or TOP10. Cells were grown at 37°C in liquid culture or on plates of Luria-Bertani (LB) medium supplemented with ampicillin (50 μg ml^{-1}).

Library construction

All libraries were constructed with QuikChange Lightning single-site- or multi-site-directed mutagenic PCR amplifications (Stratagene). Some modifications of the general protocol provided with the kit were made. DpnI digestion of parental DNA was carried out at 37°C over 40 min instead of 5 min and ultracompetent cells of strain XL10-Gold were directly transformed with *Dpn*I digested PCR product (4 μl). Generally between 2×10^3 and 8×10^3 clones were obtained on a square LB plus ampicillin plate (50 mg L^{-1}) after overnight incubation at 37°C. From the pooled colonies, plasmid DNA was isolated and stored as library.

Library pre-screening

Chemically competent *E. coli* TOP-10 cells were transformed with library plasmid DNA and plated on pH indicator plates for prescreening. Indicator

plates contained eosin (40 mg L^{-1}), methylene blue (6.5 mg L^{-1}), arabinose (200 mg L^{-1}), and ampicillin (50 mg L^{-1}).[40] Hydrobromic acid production by individual bacterial colonies could be detected after overnight growth at 37°C. For this, plates were incubated for 15 min at room temperature after spotting bromoethane (600 µl) on a filter paper that was mounted in the lid of the plates. The plates were sealed with parafilm and incubated in a fume hood. Bacterial colonies producing active haloalkane dehalogenase developed an intense purple color, and cell material was immediately transferred to 96 well plates containing LB medium (150 µl) plus ampicillin. After overnight growth at 37°C under shaking, glycerol (30 µl, 50% v:v) was added to each well and the plates were frozen at -80°C and stored at -20°C until further use.

Screening for enantioselective TCP conversion

96-Deep-well plates (polypropylene, Waters 186002482) containing LB_{amp} liquid medium (200 µL) were inoculated from frozen stocks by use of a pinned stamp (Enzyscreen). After 6 h of growth at 37°C under shaking conditions the log-phase cultures were induced by addition of LB_{amp} (1 mL) containing arabinose (0.02%) and growth was allowed to continue under the same conditions for 18 h. Pure TCP (3.5 µL) was then added directly to each culture and the plates were capped and incubated for 5 h under the same conditions. Both unreacted TCP and produced DCP were extracted from the cultures with chloroform (220 µL) containing dodecane (0.05%, v/v) as an internal standard. The organic layer was carefully transferred to phase separation plates (Whatman 96-well Unifilter 7720-7229-1) and the filtrate was collected in collection plates (Corning 3370) containing molecular sieves powder (4Å, ≈ 50 mg). Plates were sealed and incubated for 10 min under shaking conditions and the drying agent was removed by centrifugation at 2500*g* for 10 min at 16°C. The chloroform extract was transferred to polypropylene 96-deep-well plates (Greiner 780215) containing glass inserts (300 µL, Waters 150820) and plates were closed with aluminium seals.

For chiral GC we used a Shimadzu 17A dual-line GC with a PAL injection system (CTC Analytics), an FID detector, and GC-solutions software. Separation was performed with a Hydrodex-β-TBDAc column, (1 mL min^{-1}, isothermal, 138°C). Retention times were 1.83 min (TCP), 3.35 min ((*R*)-

DCP) and 3.50 min ((*S*)-DCP). Samples from one plate were injected with time intervals of 4.5 min and run in a single isothermal run, allowing measurement of 384 samples in 24 h.

Thermostability

Thermal unfolding of mutant enzymes was monitored by the fluorescence-based thermal stability method described by Ericsson *et al.*.[45] This method is based on a fluorescence increase upon binding of Sypro Orange (Life Technologies, Carlsbad, CA, USA) to hydrophobic protein surfaces that become exposed during unfolding. Increases in fluorescence emission at 575 nm were monitored by excitation at 490 nm with a MyiQ realtime PCR machine (Biorad) with increasing block temperatures from 20 to 90°C at 1°C min^{-1}. Protein solutions containing 300-fold diluted Sypro Orange (7.5 µL) in MilliQ water and purified protein (0.4 mg mL^{-1}, 17.5 µL) in TEMG buffer were prepared in iQ 96-well real-time PCR plates. After sealing with iQ 96-Well PCR plate seals (Biorad) the temperature gradient was started. The first derivative of the measured fluorescence change with temperature was calculated, giving the apparent melting temperature at the local maximum.

Circular dichroism spectroscopy was carried out with a Jasco J-815 circular dichroism spectrometer and a 3 mL quartz cuvette. Protein samples were dialyzed against potassium phosphate buffer (pH 8.2, 10 mM) and diluted to about 0.2 mg mL^{-1}. The temperature was increased at a rate of 1°C min^{-1} and spectra from 200 to 260 nm were recorded every min. Displayed spectra are averaged out of three recorded spectra. From plots of ellipticity versus temperature, the temperature at which 50% of the protein was unfolded ($T_{m,app}$) was derived by fitting a Boltzmann sigmoidal equation.

Acknowledgements

We thank Theodora Tiemersma-Wegman for her support with the automated GC-based screening assay. This project was financially supported through B-Basic, a public-private NWO-ACTS program. Part of this research was funded by the NWO (Netherlands Organization for Scientific Research) through ECHO Grant 08.B3.051. Author contributions: JGEvL performed

the library design and screening work, RJF did the thermofluor experiment, JGEvL and DBJ wrote the paper and HJW, JMvdL and DBJ corrected it.

References

1. a) F. H. Arnold, *Acc. Chem. Res.* **1998**, *31*, 125-131; b) C. A. Voigt, S. Kauffman, Z. G. Wang, *Adv. Protein Chem.* **2001**, *55*, 79-160; c) D. Bottcher, U. T. Bornscheuer, *Curr. Opin. Microbiol.* **2010**, *13*, 274-282; d) M. T. Reetz, *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174; e) A. S. Bommarius, J. K. Blum, M. J. Abrahamson, *Curr. Opin. Chem. Biol.* **2011**, *15*, 194-200.
2. a) J. C. Moore, H. M. Jin, O. Kuchner, F. H. Arnold, *J. Mol. Biol.* **1997**, 272, 336-347; b) L. Giver, A. Gershenson, P. O. Freskgard, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12809-12813; c) K. Miyazaki, P. L. Wintrode, R. A. Grayling, D. N. Rubingh, F. H. Arnold, *J. Mol. Biol.* **2000**, *297*, 1015-1026; d) S. Ahmad, M. Z. Kamal, R. Sankaranarayanan, N. M. Rao, *J. Mol. Biol.* **2008**, *381*, 324-340; e) J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458-467; T. S. Wong, F. H. Arnold, U. Schwaneberg, *Biotechnol. Bioeng.* **2004**, *85*, 351-358; f) L. Zhu, K. L. Tee, D. Roccatano, B. Sonmez, Y. Ni, Z. H. Sun, U. Schwaneberg, *ChemBioChem* **2010**, *11*, 691-697; g) Y. Qin, X. Wei, X. Song, Y. Qu, *J. Biotechnol.* **2008**, *135*, 190-195.
3. a) J. H. Zhang, G. Dawes, W. P. Stemmer, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4504-4509; b) N. J. Turner, Trends *Biotechnol.* **2003**, *21*, 474-478; c) P. A. Dalby, *Curr. Opin. Struct. Biol.* **2003**, *13*, 500-505; T. Sakamoto, J. M. Joern, A. Arisawa, F. H. Arnold, *Appl. Environ. Microbiol.* **2001**, *67*, 3882-3887.
4. a) M. T. Reetz, *J. Org. Chem.* **2009**, *74*, 5767-5778; b) Y. L. Boersma, T. Pijning, M. S. Bosma, A. M. van der Sloot, L. F. Godinho, M. J. Drçge, R. T. Winter, G. van Pouderoyen, B. W. Dijkstra, W. J. Quax, *Chem. Biol.* **2008**, *15*, 782-789; c) A. Kirschner, U. T. Bornscheuer, *Appl. Microbiol. Biotechnol.* **2008**, *81*, 465-472; d) M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5716-5722; e) M. Schmidt, D. Hasenpusch, M. K_hler, U. Kirchner, K. Wiggenhorn, W. Langel, U. T. Bornscheuer, *ChemBioChem* **2006**, *7*, 805-809.
5. a) R. J. Kazlauskas, U. T. Bornscheuer, *Nat. Chem. Biol.* **2009**, *5*, 526-529; b) H. Jochens, U. T. Bornscheuer, *ChemBioChem* **2010**, *11*, 1861-1866; c) M. J. Dougherty, F. H. Arnold, *Curr. Opin. Biotechnol.* **2009**, *20*, 486-491; d) M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem* **2008**, *9*, 1797-1804; e) J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. Mac-Millan, F. H. Arnold, *Curr. Opin. Struct. Biol.* **2005**, *15*, 447-452; f) S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291-297.
6. a) R. Cadwell, G. Joyce, *PCR Meth. Appl.* **1992**, *2*, 28-33; b)W. P. Stemmer, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10747-10751; c) A. M. Aguinaldo, F. H. Arnold, *Mol. Biol.* **2003**, *231*, 105-110; d) T. S. Wong, K. L. Tee, B. Hauer, U. Schwaneberg, *Nucleic Acids Res.* **2004**, *32*, e26.
7. a) M. T. Reetz, J. D. Carballeira, J. Peyralans, H. Hçbenreich, A. Maichele, A. Vogel, *Chem. Eur. J.* **2006**, *12*, 6031-6038; b) M. T. Reetz, S. Prasad, J. D.

Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.* **2010**, *132*, 9144-9152; c) M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891-903.

8. a) A. Pavelka, E. Chovancova, J. Damborsky, *Nucleic Acids Res.* **2009**, *37*, W376-383; b) R. Kourist, H. Jochens, S. Bartsch, R. Kuipers, S. K. Padhi, M. Gall, D. Bçttcher, H. J. Joosten, U. T. Bornscheuer, *ChemBioChem* **2010**, *11*, 1635-1643.
9. M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem.* **2006**, *118*, 7909-7915; *Angew. Chem. Int. Ed.* **2006**, *45*, 7745-7751.
10. a) P. Heinzelman, C. D. Snow, M. A. Smith, X. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull, F. H. Arnold, *J. Biol. Chem.* **2009**, *284,* 26229-26233; b) P. Heinzelman, R. Komor, A. Kanaan, P. Romero, X. Yu, S. Mohler, C. Snow, F. H. Arnold, *Protein Eng. Des. Sel.* **2010**, *23*, 871-880; c) C. R. Otey, M. Landwehr, J. B. Endelman, K. Hiraga, J. D. Bloom, F. H. Arnold, *PLoS Biol.* **2006**, *4*, e112; d) D. Zha, A. Eipper, M. T. Reetz, *ChemBioChem* **2003**, *4*, 34-39.
11. a) J. van den Brulle, M. Fischer, T. Langmann, G. Horn, T. Waldmann, S. Arnold, M. Fuhrmann, O. Schatz, T. O'Connell, D. O'Connell, A. Auckenthaler, H. Schwer, *Biotechniques* **2008**, *45*, 340-343; b) T. S. Wong, D. Roccatano, D. Loakes, K. L. Tee, A. Schenk, B. Hauer, U. Schwaneberg, *J. Biotechnol.* **2008**, *3*, 74-82.
12. a) M. T. Reetz, S. Wu, *Chem. Commun.* **2008**, *43*, 5499-5501; b) E. Munoz, M. W. Deem, *Protein Eng. Des. Sel.* **2009**, *21*, 311-317; c) K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa, A. S. Bommarius, *J. Biotechnol.* **2006**, *1*, 531-536.
13. C. J_ckel, J. D. Bloom, P. Kast, F. H. Arnold, D. Hilvert, *J. Mol. Biol.* **2010**, *399*, 541-546.
14. a) R. K. Kuipers, H. J. Joosten, W. J. van Berkel, N. G. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. dos Santos, P. J. Schaap, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 2101-2113; b) H. Jochens, D. Aerts, U. T. Bornscheuer, *Protein Eng. Des. Sel.* **2010**, *23*, 903-909.
15. a) J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *J. Biotechnol.* **2007**, *2*, 180-191; b) R. J. Hayes, J. Bentzien, M. L. Ary, M. Y. Hwang, J. M. Jacinto, J. Vielmetter, A. Kundu, B. I. Dahiyat, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15926-15931; c) L. G. Otten, F. Hollmann, I. W. Arends, *Trends Biotechnol.* **2010**, *28*, 46-54; d) M. T. Reetz, P. Soni, L. Fernandez, *Biotechnol. Bioeng.* **2009**, *102*, 1712-1717.
16. M. A. Mena, P. S. Daugherty, *Protein Eng. Des. Sel.* **2005**, *18*, 559-561.
17. R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338-344.
18. P. F. Mugford, U. G. Wagner, Y. Jiang, K. Faber, R. J. Kazlauskas, *Angew. Chem.* **2008**, *120*, 8912-8923; *Angew. Chem. Int. Ed.* **2008**, *47*, 8782-8793.
19. a) C. M. Clouthier, M. M. Kayser, M. T. Reetz, *J. Org. Chem.* **2006**, *71*, 8431-8437; b) M. T. Reetz, F. Daligault, B. Brunner, H. Hinrichs, A. Deege, *Angew. Chem.* **2004**, *116*, 4170-4173; *Angew. Chem. Int. Ed.* **2004**, *43*, 4078-4081.
20. S. F. Royer, L. Haslett, S. J. Crennell, D. W. Hough, M. J. Danson, S. D. Bull, *J. Am. Chem. Soc.* **2010**, *132*, 11753-11758.

21. a) S. K. Padhi, D. J. Bougioukou, J. D. Stewart, *J. Am. Chem. Soc.* **2009**, *131*, 3271-3280; b) D. J. Bougioukou, S. Kille, A. Taglievber, M. T. Reetz, *Adv. Synth. Catal.* **2009**, *351*, 3287-3305.
22. T. Kubo, M. W. Peters, P. Meinhold, F. H. Arnold, *Chem. Eur. J.* **2006**, *12*, 1216-1220.
23. H. Zheng, M. T. Reetz, *J. Am. Chem. Soc.* **2010**, *132*, 15744-15751.
24. J. W. Bijsterbosch, A. Das, F. P. J. M. Kerkhof, *J. Cleaner Prod.* **1994**, *2*, 181-184.
25. a) T. Bosma, J. Damborsky´, G. Stucki, D. B. Janssen, *Appl. Environ. Microbiol.* **2002**, *68*, 3582-3587; b) K. A. Gray, T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, R. Knowles, L. Kan, P. E. Swanson, D. E. Robertson, *Adv. Synth. Catal.* **2001**, *343*, 607-617; c) M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727-733.
26. a) N. Kasai, K. Sakaguchi, *Tetrahedron Lett.* **1992**, *33*, 1211-1212; b) N. Kasai, K. Sakaguchi, *Adv. Synth. Catal.* **2003**, *345*, 437-455.
27. a) P. Popik, M. Krawczyk, K. Golembiowska, G. Nowak, A. Janowsky, P. Skolnick, A. Lippa, A. S. Basile, *Cell. Mol. Neurobiol.* **2006**, *26*, 855-871; b) M. M. Kabat, A. R. Daniewski, W. Burger, *Tetrahedron: Asymmetry* **2007**, *18*, 2663-2665.
28. A. M. Thayer, *Chem. Eng. News* **2006**, *84*, 26-27.
29. M. Breuning, M. Winnacker, M. Steiner, *Eur. J. Org. Chem.* **2007**, 2100-2106.
30. a) D. B. Janssen, *Curr. Opin. Chem. Biol.* **2004**, *8*, 150-159; b) E. Chovancov _, J. Kosinski, J.M. Bujnicki, J. Damborsky, *Proteins* **2007**, *67*, 305-316.
31. a) A. N. Kulakova, M. J. Larkin, L. A. Kulakov, *Microbiology* **1997**, *143*, 109-115; b) G. J. Poelarends, M. Zandstra, T. Bosma, L. A. Kulakov, M. J. Larkin, J. R. Marchesi, A. J. Weightman, D. B. Janssen, *J. Bacteriol.* **2000**, *182*, 2725-2731; c) G. J. Poelarends, L. A. Kulakov, M. J. Larkin, J. E. van Hylckama Vlieg, D. B. Janssen, *J. Bacteriol.* **2000**, *182*, 2191-2199; d) D. B. Janssen, I. J. Dinkla, G. J. Poelarends, P. Terpstra, *Environ. Microbiol.* **2005**, *7*, 1868-8182.
32. a) Z. Prokop, Y. Sato, J. Brezovsky, T. Mozga, R. Chaloupkova, T. Koudelakova, P. Jerabek, V. Stepankova, R. Natsume, J. G. van Leeuwen, D. B. Janssen, J. Florian, Y. Nagata, T. Senda, J. Damborsky, *Angew. Chem. Int. Ed.* **2010**, *49*, 6111-6115; b) A. Westerbeek, W. Szymanski, H. J. Wijma, S. J. Marrink, B. L. Feringa, D. B. Janssen, *Adv. Synth. Catal.* **2011**, *353*, 931-944; c) R. J. Pieters, J. H. Lutje Spelberg, R. M. Kellogg, D. B. Janssen, *Tetrahedron Lett.* **2001**, *42*, 469-471.
33. M. T. Reetz, *Angew. Chem.* **2002**, *114*, 1391-1394; *Angew. Chem. Int. Ed.* **2002**, *41*, 1335-1338.
34. J. Newman, T. S. Peat, R. Richard, L. Kan, P. E. Swanson, J. A. Affholter, I. H. Holmes, J. F. Schindler, C. J. Unkefer, T. C. Terwilliger*, Biochemistry* **1999**, *38*, 16105-16114.
35. W. M. Patrick, A. E. Firth, J. M. Blackburn, *Protein Eng.* **2003**, *16*, 451-457
36. a) M. T. Reetz, M. Puls, J. D. Carballeira, A. Vogel, K. E. Jaeger, T. Eggert, W. Thiel, M. Bocola, N. Otte, *ChemBioChem* **2007**, *8*, 106-112; b) M. T. Reetz, P. Soni, J. P. Acevedo, J. Sanchis, *Angew. Chem.* **2009**, *121*, 8418-8422; *Angew. Chem. Int. Ed.* **2009**, *48*, 8268-8272.

37. a) G. P. Horsman, A. M. F. Liu, E. Henke, U. T. Bornscheuer, R. J. Kaslauskas, *J. Chem. Eur.* **2003**, *9*, 1933-1939; b) S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol* **2005**, *12*, 45-54.
38. a) M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260-2267; b) C. A. Tracewell, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2009**, *13*, 3-9.
39. F. Pries, A. J. van den Wijngaard, R. Bos, M. Pentenga, D. B. Janssen, *J. Biol. Chem.* **2004**, *269*, 17490-17494.
40. M. A. Loos, *J. Can. Microbiol.* **1975**, *21*, 104-107.
41. a) E. Krieger, K. Joo, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus, *Proteins Struct. Funct. Bioinf.* **2009**, *77*, 114-122; b) E. Krieger, G. Koraimann, G. Vriend, *Proteins Struct. Funct. Bioinf.* **2002**, *47*, 393-402.
42. a) A. Stsiapanava, J. Dohnalek, J. A. Gavira, M. Kuty, T. Koudelakova, J. Damborsky, Kuta, I. Smatanova, Crystallogr. *Biol. Crystallogr. D.* **2010**, *66*, 962-969.
43. a) G. M. Morris, D. S. Goodsell, R. S. Haliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639-1662; b) G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, *30*, 2785-2791.
44. R. M. de Jong, J. J. Tiesinga, H. J. Rozeboom, K. H. Kalk, L. Tang, D. B. Janssen, B. W. Dijkstra, *EMBO J.* **2003**, *22*, 4933-4944.
45. U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, P. Nordlund, *Anal. Biochem.* **2006**, *357*, 289-298.